

## PATENT ABSTRACTS OF JAPAN

6

(11)Publication number : 07-056694

(43)Date of publication of application : 03.03.1995

(51)Int.Cl. G06F 3/06  
 G06F 3/06  
 G06F 11/16  
 G06F 12/16

(21)Application number : 05-205430

(71)Applicant : FUJI XEROX CO LTD

(22)Date of filing : 19.08.1993

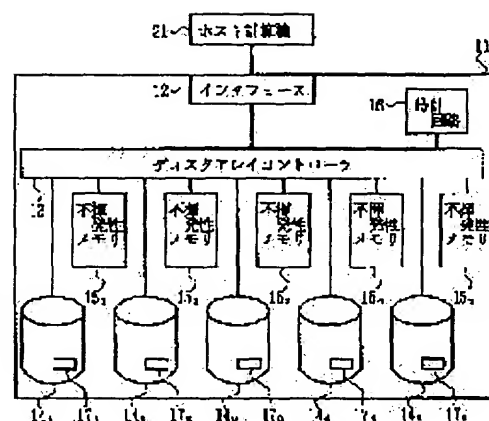
(72)Inventor : TSUTSUMI KENJI

## (54) DISK ARRAY DEVICE

## (57)Abstract:

PURPOSE: To provide a disk array device in which the loss of data due to an artificial miss can be prevented from being generated by exactly detecting the state of each incorporated disk at the time of rerising.

CONSTITUTION: A mechanism which compares management information written in the prescribed storage area of each disk device with management information in a non-volatile memory at the time of rerising, and judges the disk device as the family disk device when they are not matched, and a mechanism which rewrites the content of the management information in the non-volatile memory corresponding to the disk device in which a fault occurs are added to a disk array controller 13. Thus, after an abnormal situation occurs, or after the disk device is exchanged at the time of system stop, the management information in the storage area is different from the management information in the non-volatile memory, and the state of each disk device can be exactly grasped based on those management information in a disk array device 11.



## LEGAL STATUS

[Date of request for examination]

[Date of sending the examiner's decision of rejection]

[Kind of final disposal of application other than the examiner's decision of rejection or application converted registration]

[Date of final disposal for application]

[Patent number]

[Date of registration]

[Number of appeal against examiner's decision of rejection]

[Date of requesting appeal against examiner's decision of rejection]

[Date of extinction of right]

(19)日本国特許庁 (J P)

(12) 公開特許公報 (A)

(11)特許出願公開番号

特開平7-56694

(43)公開日 平成7年(1995)3月3日

(51)Int.Cl. <sup>6</sup>	識別記号	庁内整理番号	F I	技術表示箇所
G 0 6 F 3/06	5 4 0	7165-5B		
	3 0 4 F			
11/16	3 1 0 H			
12/16	3 1 0 A	9293-5B		

審査請求 未請求 請求項の数 3 O L (全 11 頁)

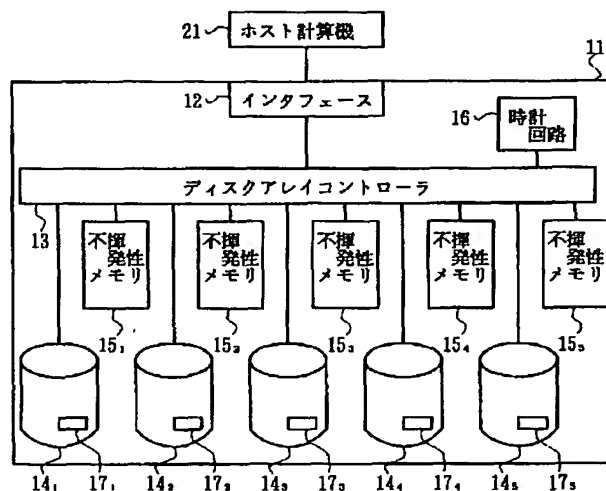
(21)出願番号	特願平5-205430	(71)出願人	000005496 富士ゼロックス株式会社 東京都港区赤坂三丁目3番5号
(22)出願日	平成5年(1993)8月19日	(72)発明者	堤 健次 埼玉県岩槻市府内3丁目7番1号 富士ゼロックス株式会社内
		(74)代理人	弁理士 山内 梅雄

(54)【発明の名称】 ディスクアレイ装置

(57)【要約】

【目的】 内蔵する各ディスク装置の状態を、再立ち上げ時に正確に検出して、人為的ミスによるデータの喪失が生じることのないディスクアレイ装置を提供する。

【構成】 再立ち上げが行われた際に、各ディスク装置14の所定の記録領域17に書き込まれている管理情報と不揮発性メモリ15内の管理情報とを比較し、それらが一致していない場合には、そのディスク装置を障害ディスク装置とする機構と、内蔵するディスク装置14のいずれかに障害が発生した際に、そのディスク装置に対応する不揮発性メモリ15の管理情報の内容の書き換えを行う機構をディスクアレイコントローラ13に付加する。これにより、異常事態発生後や、システム停止時にディスク装置を交換した後は、記録領域17と不揮発性メモリ15内の管理情報が異なるようになり、ディスクアレイ装置11は、これらの管理情報を基に、各ディスク装置の状態を正確に把握することができる。



1

## 【特許請求の範囲】

【請求項 1】 データを冗長構成にして格納するための複数のディスク装置と、  
前記複数のディスク装置のそれぞれの状態の管理に用いる管理情報を格納するための不揮発性メモリと、  
初期化が指示されたときに、それぞれのディスク装置の所定の記録領域に、同一の管理情報を書き込むとともに、それらの管理情報をそれぞれのディスク装置に対応させて前記不揮発性メモリに書き込む管理情報書込手段と、  
前記複数のディスク装置のいずれかに読み出したまたは書き込み障害が発生したことを検出した際に、そのディスク装置に対応する前記不揮発性メモリ内の管理情報の内容を書き換える管理情報書換手段と、  
再立ち上げが行われたときに、各ディスク装置の前記記録領域に書き込まれている管理情報と、そのディスク装置に対応する不揮発性メモリ内に格納されている管理情報とを比較し、それらが一致していないディスク装置を障害ディスク装置として処理するディスク装置処理手段とを具備することを特徴とするディスクアレイ装置。

【請求項 2】 データを冗長構成にして格納するための複数のディスク装置と、  
前記複数のディスク装置のそれぞれの状態の管理に用いる管理情報を格納するための不揮発性メモリと、  
初期化が指示されたときに、それぞれのディスク装置の所定の記録領域に、互いに異なる管理情報を書き込むとともに、それらの管理情報を各ディスク装置に対応させて前記不揮発性メモリに書き込む管理情報書込手段と、  
前記複数のディスク装置のいずれかに読み出したまたは書き込み障害が発生したことを検出した際に、そのディスク装置に対応する前記不揮発性メモリ内の管理情報を、他の不揮発性メモリ内管理情報と一致しない管理情報で書き換える管理情報書換手段と、  
再立ち上げが行われたときに、各ディスク装置の前記記録領域に書き込まれている管理情報と、そのディスク装置に対応する不揮発性メモリ内に格納されている管理情報とを比較し、それらが一致していないディスク装置を障害ディスク装置として処理するディスク装置処理手段とを具備することを特徴とするディスクアレイ装置。

【請求項 3】 データを冗長構成にして格納するための複数のディスク装置と、  
前記複数のディスク装置のそれぞれの状態の管理に用いる管理情報を格納するための不揮発性メモリと、  
初期化が指示されたときに、それぞれのディスク装置の所定の記録領域に、互いに異なる管理情報を書き込むとともに、それらの管理情報を各ディスク装置に対応させて前記不揮発性メモリに書き込む管理情報書込手段と、  
前記複数のディスク装置のいずれかに読み出したまたは書き込み障害が発生したことを検出した際に、そのディスク装置に対応する前記不揮発性メモリ内の管理情報を、

2

他の不揮発性メモリ内管理情報と一致しない管理情報で書き換える管理情報書換手段と、  
再立ち上げが行われたときに、各ディスク装置の前記記録領域に書き込まれている管理情報と、そのディスク装置に対応する不揮発性メモリ内に格納されている管理情報とを比較し、それらが一致していないディスク装置を特定するディスク装置特定手段と、  
この管理情報不一致ディスク装置特定手段により特定されたディスク装置の動作チェックを行い、そのディスク装置が正常にアクセスが行えない場合には、そのディスク装置を障害ディスク装置として処理し、正常にアクセスが行える場合には、他のディスク装置内に格納されているデータを基にそのディスク装置を復旧させ、復旧させたディスク装置の所定の記録領域に、対応する前記不揮発性メモリ内の管理情報を書き込むディスク装置処理手段とを具備することを特徴とするディスクアレイ装置。

## 【発明の詳細な説明】

## 【0001】

【産業上の利用分野】 本発明は、ディスクアレイ装置に係わり、特に、データを冗長構成にして管理するディスクアレイ装置に関する。

## 【0002】

【従来の技術】 ディスクアレイ装置とは、複数のディスク装置を組み合わせ、ホスト側から 1 つのディスク装置として扱えるようにした記憶装置である。この装置では、データを複数のディスク装置に分散して格納できるため、単体のディスク装置を凌駕するアクセス速度を実現し得る。通常、ディスクアレイ装置では、データの書き込み時に、パリティ等のエラー訂正符号の算出を行い、それらを複数のディスク装置に分散して書き込むことにより、あるディスク装置に格納されているデータが、他のディスク装置に格納されているデータから再構築できるようにする。このようにデータを冗長構成とすることにより、内蔵するディスク装置の 1 台に格納していたデータが失われても、ディスクアレイ装置全体としてみた場合、そのデータは失われていないことになり、記憶装置としての信頼性を向上させることができる。このため、大容量の外部記憶装置が必要とされるワークステーションなどでは、ディスクアレイ装置が用いられるようになってきている。

【0003】 図 5 に、従来のディスクアレイ装置の構成を示す。ディスクアレイ装置 11 は、インタフェース 12 とディスクアレイコントローラ 13 と複数台のディスク装置 141 ～ 145 で構成される。インタフェース 12 は、ホスト計算機 21 とディスクアレイコントローラ 13 に接続され、両者の間のデータ転送を受け持つ。ディスクアレイコントローラ 13 は、インタフェース 12 を介して入力されるホスト計算機 21 からのアクセス要求に応じて、各ディスク装置 141 ～ 145 を制御し

て、データの書き込みと読み出しを行う制御回路であり、制御用プロセッサとプログラム用メモリとバッファとして用いるデータ用メモリと演算回路で構成される。

【0004】図6を用いて、RAID(Redundant Array of Inexpensive Disks)のレベル4と呼ばれる冗長構成を用いた場合のディスクアレイ装置の動作を簡単に説明する。ディスクアレイコントローラ13は、ホスト計算機からの書き込み要求がなされると、書き込み要求がなされたデータ20を、セクタ単位に分割し、ディスク装置141～144の記録領域181～184に分散させて書き込む。このとき、ディスクアレイコントローラ13は、単にデータを分散するだけでなく、対応する記録領域181～184に格納されるデータD1、D2、D3、D4の排他的論理和演算を行い、その演算結果であるパリティPをディスク装置145の記録領域185に書き込む。

【0005】このような書き込み動作により、データに冗長性が付与され、いずれのデータも他のデータ、パリティを基に再構築可能な形態で格納されることになる。たとえば、データD4は、データD1、D2、D3とパリティPの排他的論理和演算結果と等しくなっており、記録領域184が読み出し不可能となった場合には、他のディスク装置の記録領域のデータ、パリティを読み出して、これらの排他的論理和演算を行うことにより、記録領域184を読み出すことなく、データD4を得ることができる。

【0006】このような動作においては、アクセス要求のあったディスク装置が、読み出し不可能なディスク装置(障害ディスク装置)であるか否かの判断を行う必要がある。従来のディスクアレイ装置では、ディスクアレイコントローラ13の内部のメモリにその障害ディスク装置を特定する情報を記憶しておき、その情報を基に再構築による応答か否かを選択していた。

【0007】しかし、障害ディスク装置を特定する情報が格納されるメモリが揮発性メモリである場合には、停電等によるシステムダウンが発生したときに、障害ディスク装置を特定する情報が失われてしまい、再立ち上げ時にその情報を使用者が再設定しなければならないという問題があった。このため、障害ディスク装置を特定する情報を揮発性メモリ以外に格納する装置が各種提案されている。

【0008】たとえば、特開平3-229331号公報には、2重化ディスク装置の状態記憶に不揮発性メモリを用いるシステムが提案されている。図7に、その概要を示す。このシステムは、2台のディスク装置22と制御部23と不揮発性記憶部24とオペレーティングシステム(OS)25で構成される。制御部23は、いずれかのディスク装置22に読み出し、または、書き込み障害が発生したことを検出すると、その障害ディスク装置とは異なる方のディスク装置22の装置機番を不揮発性

記憶部24に書き込む。システムの立ち上げ時には、制御部23は、不揮発性記憶部24内の装置機番を読み出すことにより、正常なディスク装置の識別を行う。

【0009】また、特公平2-32652号公報には、2重化ディスク装置においてディスク媒体上に記録された情報から、ディスク装置の状態を判定するシステムが提案されている。図8に、その概要を示す。このシステムでは、2台のディスク装置221と222は、それぞれチャネル装置261と262により中央処理装置27に接続され、各ディスク装置のディスク媒体上には、ヒストリカウンタ部29とヒストリカウンタ検証部30が設けられ、中央処理装置27内の主メモリ28には、制御テーブル31が設けられている。この制御テーブル31は、各ディスク装置がオンライン状態にあるかオフライン状態にあるかを示す情報を保持するために用いられる。

【0010】このシステムでは、両方のディスク装置22が、正常に機能しているときに、それぞれのヒストリカウンタ部29に同じ値を書き込んでおく。そして、一方のディスク装置、たとえば、ディスク装置221がシステムから切り離された場合には、他方のディスク媒体上のヒストリカウンタ部292の値を増加させる。システムが再立ち上げされたときには、2つのヒストリカウンタ部291と292の値を読み出して比較し、それらの値が同一の値である場合には、両方のディスク装置がオンライン状態にあったと判断して、その旨を制御テーブル31に書き込む。また、それらの値が異なるときには、大きな値が書き込まれている方のディスク装置がオンライン状態にあったと判断し、そのディスク装置がオンライン状態となり、他方のディスク装置がオフライン状態となるように制御テーブルの内容を設定する。なお、ヒストリカウンタ検証部30は、ヒストリカウンタ部29に隣接して設けられた記録領域であり、この装置では、その部分に書き込まれたデータの変更の有無を確認することにより、ヒストリカウンタ部の内容の正否の判定を行っている。

【0011】

【発明が解決しようとする課題】前述した不揮発性記憶部を用いたシステムによれば、障害が発生したディスク装置を特定する情報が停電等を原因としたシステムダウンにより失われることはない。しかし、障害が発生した際に、そのディスク装置を特定する情報を不揮発性のメモリに格納する手段を設けるだけでは、再立ち上げ時に誤動作が生じてしまうことがある。たとえば、図5に示したディスクアレイ装置に不揮発性メモリを設けた場合を考える。このディスクアレイ装置において、ディスク装置144に障害が発生したときには、不揮発性メモリに、その障害ディスク装置を特定する情報が格納される。この障害ディスク装置は、システム停止時などに交換されるわけだが、この際に、交換を行う作業者が誤っ

たディスク装置（たとえば、ディスク装置142）を、交換してしまい、その誤接続に気付かずディスク装置144（または、ディスク装置142）の復旧開始を指示したときには、ディスク装置内のデータの一部が失われてしまうことになる。

【0012】また、ディスク媒体上にヒストリカウンタ部を設けた2重化ディスク装置では、再立ち上げ時に、2つのディスク媒体のヒストリカウンタ部の値を比較することにより、システム停止時のディスク装置の交換も認識できる。しかし、このディスク状態の判定方法は2重化ディスク装置に対するものであるため、3つ以上のディスク装置を内蔵するディスクアレイ装置には、直接、適用することはできない。

【0013】たとえば、2重化ディスク装置では、2つのディスク装置のディスク媒体上に記憶されているデータは、通常、同一であるため、2つのディスク装置を入れ替えても、正常に動作させることは可能である。しかし、3つ以上のディスク装置を用いるディスクアレイ装置では、それぞれのディスク装置のディスク媒体に書き込まれている内容は、図6に示したように、互いに異なっている。このため、ディスクアレイ装置に内蔵されているディスク装置のうち、2台のディスク装置を入れ替えた場合、正常なデータのアクセスが行えなくなる。

【0014】このような問題は、ディスク装置の接続作業と再立ち上げ時の情報設定作業を、ディスクアレイ装置の使用者が間違いないく行えば、生じないものである。しかし、ディスクアレイ装置がデータを冗長構成とすることにより、記憶装置としての信頼性を向上させた装置である以上、このような人為的なミスにも対応し得る装置であることが望まれる。

【0015】そこで本発明の目的は、内蔵する各ディスク装置の状態を、常に正確に把握して、人為的ミスによるデータの喪失が生じることのないディスクアレイ装置を提供することにある。

【0016】

【課題を解決するための手段】請求項1記載の発明では、データを冗長構成にして格納するための複数のディスク装置と、複数のディスク装置のそれぞれの状態の管理に用いる管理情報を格納するための不揮発性メモリと、初期化が指示されたときに、それぞれのディスク装置の所定の記録領域に、同一の管理情報を書き込むとともに、それらの管理情報をそれぞれのディスク装置に対応させて不揮発性メモリに書き込む管理情報書込手段と、複数のディスク装置のいずれかに読み出したまたは書き込み障害が発生したことを検出した際に、そのディスク装置に対応する不揮発性メモリ内の管理情報の内容を書き換える管理情報書換手段と、再立ち上げが行われたときに、各ディスク装置の記録領域に書き込まれている管理情報と、そのディスク装置に対応する不揮発性メモリ内に格納されている管理情報とを比較し、それらが一

致していないディスク装置を障害ディスク装置として処理するディスク装置処理手段とを具備する。

【0017】すなわち請求項1記載の発明では、自装置の再立ち上げが行われた際に、各ディスク装置のディスク媒体上に設けられた所定の記録領域に書き込まれている管理情報と不揮発性メモリ内の管理情報とを比較し、それらの管理情報が一致していない場合には、そのディスク装置を障害ディスク装置とする。管理情報書込手段は、ディスクアレイ装置の初期化時に、各ディスク装置の所定の記録領域に、同一の管理情報、たとえば、所定の数値を書き込む。これらの管理情報は、不揮発性メモリ内にも、それぞれのディスク装置に対応させて書き込まれる。管理情報書換手段は、内蔵するディスク装置のいずれかに読み出したまたは書き込み障害が発生した際に、そのディスク装置に対応する不揮発性メモリの管理情報の内容の書き換えを行う。これらにより、障害が発生したディスク装置では、その記録領域に書き込まれた管理情報と、対応する不揮発性メモリ内の管理情報とが異なるようになるため、内蔵する各ディスク装置の状態を、常に正確に把握することができる。

【0018】請求項2記載の発明では、データを冗長構成にして格納するための複数のディスク装置と、複数のディスク装置のそれぞれの状態の管理に用いる管理情報を格納するための不揮発性メモリと、初期化が指示されたときに、それぞれのディスク装置の所定の記録領域に、互いに異なる管理情報を書き込むとともに、それらの管理情報を各ディスク装置に対応させて不揮発性メモリに書き込む管理情報書込手段と、複数のディスク装置のいずれかに読み出したまたは書き込み障害が発生したことを検出した際に、そのディスク装置に対応する不揮発性メモリ内の管理情報を、他の不揮発性メモリ内管理情報と一致しない管理情報で書き換える管理情報書換手段と、再立ち上げが行われたときに、各ディスク装置の前記記録領域に書き込まれている管理情報と、そのディスク装置に対応する不揮発性メモリ内に格納されている管理情報とを比較し、それらが一致していないディスク装置を障害ディスク装置として処理するディスク装置処理手段とを具備する。

【0019】すなわち請求項2記載の発明では、自装置の再立ち上げが行われた際に、各ディスク装置のディスク媒体上に設けられた所定の記録領域に書き込まれている管理情報と不揮発性メモリ内の管理情報とを比較し、それらの管理情報が一致していない場合には、そのディスク装置を障害ディスク装置とする。管理情報書込手段は、ディスクアレイ装置の初期化時に、各ディスク装置の所定の記録領域に、互いに内容の異なる管理情報を書き込む。これらの管理情報は、不揮発性メモリ内にも、それぞれのディスク装置に対応させて書き込まれる。管理情報としては、たとえば、乱数発生器により発生させた乱数や、ディスク装置のスロット番号や、日付と時間

を基に作成した情報などを用いることができる。

【0020】管理情報書換手段は、内蔵するディスク装置のいずれかに読み出したり書き込み障害が発生した際に、そのディスク装置に対応する不揮発性メモリの管理情報の内容を他の不揮発性メモリ内管理情報と一致しない管理情報で書き換えを行う。管理情報書換手段で管理情報としてディスク装置のスロット番号を書き込んだ場合には、たとえば、障害が発生するたびに、管理情報をディスク装置の台数分増加させていけば、変更後の管理情報が他の管理情報と一致しないようにすることができる。これらにより、読み出したり書き込み障害が発生したディスク装置では、その記録領域に書き込まれた管理情報と、対応する不揮発性メモリ内の管理情報とが異なるようになるため、内蔵する各ディスク装置の状態を、常に正確に把握することができる。

【0021】請求項3記載の発明では、データを冗長構成にして格納するための複数のディスク装置と、複数のディスク装置のそれぞれの状態の管理に用いる管理情報を格納するための不揮発性メモリと、初期化が指示されたときに、それぞれのディスク装置の所定の記録領域に、互いに異なる管理情報を書き込むとともに、それらの管理情報を各ディスク装置に対応させて不揮発性メモリに書き込む管理情報書換手段と、複数のディスク装置のいずれかに読み出したり書き込み障害が発生したことを検出した際に、そのディスク装置に対応する前記不揮発性メモリ内の管理情報を、他の不揮発性メモリ内管理情報と一致しない管理情報で書き換える管理情報書換手段と、再立ち上げが行われたときに、各ディスク装置の記録領域に書き込まれている管理情報と、そのディスク装置に対応する不揮発性メモリ内に格納されている管理情報とを比較し、それらが一致していないディスク装置を特定するディスク装置特定手段と、このディスク装置特定手段により特定されたディスク装置の動作チェックを行い、そのディスク装置が正常にアクセスが行えない場合には、そのディスク装置を障害ディスク装置として処理し、正常にアクセスが行える場合には、他のディスク装置内に格納されているデータを基にそのディスク装置を復旧させ、復旧させたディスク装置の所定の記録領域に、対応する不揮発性メモリ内の管理情報を書き込むディスク装置処理手段とを具備する。

【0022】すなわち請求項3記載の発明では、ディスク装置処理手段により、自装置の再立ち上げが行われた際に、各ディスク装置のディスク媒体上に設けられた所定の記録領域に書き込まれている管理情報と不揮発性メモリ内の管理情報とが比較され、それらの管理情報が一致していない場合には、そのディスク装置の動作チェックが行われ、そのディスク装置が、アクセス可能であった場合には、他のディスク装置のデータからそのディスク装置が復旧され、復旧完了後、そのディスク媒体の所定の記憶領域に、対応する管理情報を書き込まれる。管

理情報書換手段は、ディスクアレイ装置の初期化時に、各ディスク装置の所定の記録領域に、互いに内容の異なる管理情報を書き込む。これらの管理情報は、不揮発性メモリ内にも、それぞれのディスク装置に対応させて書き込まれる。管理情報書換手段は、内蔵するディスク装置のいずれかに読み出したり書き込み障害が発生した際に、そのディスク装置に対応する不揮発性メモリの管理情報の内容を他の不揮発性メモリ内管理情報と一致しない管理情報で書き換えを行う。これらにより、読み出したり書き込み障害が発生したディスク装置では、その記録領域に書き込まれた管理情報と、対応する不揮発性メモリ内の管理情報とが異なるようになるため、内蔵する各ディスク装置の状態を、常に正確に把握することができる。

【0023】

【実施例】以下、実施例につき本発明を詳細に説明する。

【0024】図1に、本発明の実施例によるディスクアレイ装置の構成を示す。ディスクアレイ装置11は、インタフェース12とディスクアレイコントローラ13と5台のディスク装置14<sub>1</sub>～14<sub>5</sub>と各ディスク装置に対応した不揮発性メモリ15<sub>1</sub>～15<sub>5</sub>と時計回路16とを備える。ホスト計算機21からのデータのアクセス要求(読み出したり書き込み要求)は、インタフェース12を介してディスクアレイコントローラ13に入力される。ディスクアレイコントローラ13は、その要求内容に応じて、ディスク装置14を制御して、データの読み出したり書き込みを行う機構と、不揮発性メモリ15内の情報とそれぞれのディスク媒体上に設けられた管理情報記録領域17に格納された情報を基に、各ディスク装置の状態の判定を行う機構を有する。時計回路16は、詳細は後述する管理情報の作成に用いられる。

【0025】前者の機構は、従来のディスクアレイコントローラと同様のものであるため、その動作の説明は省略し、後者の機構の動作を中心に実施例のディスクアレイ装置の説明を行う。なお、この機構は、制御用プロセッサとその制御用プロセッサの動作を規定するプログラムが格納されたプログラムメモリと動作時に使用するパラメータを格納するメモリと各ディスク装置14に読み出したり書き込み要求を行うためのインターフェイスと各不揮発性メモリ15へのインターフェイスとで構成されている。まず、初期化時の動作を説明する。

【0026】図2に、実施例のディスクアレイ装置の初期化時の動作の流れを示す。この動作は、ホスト計算機からの初期化指示により開始される。初期化指示を受信したディスクアレイコントローラ13は、カウンタ*i*に“1”、障害ディスク装置台数カウンタ*NERR*に“0”、障害ディスク装置識別パラメータ*Noisk*に“0”をセット(ステップS101)する。カウンタ*i*は、ディスク装置を識別するためのカウンタであり、障

害ディスク装置台数カウンタNERRは、使用不可能なディスク装置の台数のカウントに用いられ、障害ディスク装置識別パラメータNdiskには、使用不可とするディスク装置のスロット番号が格納される。これらのパラメータの格納には、ディスクアレイコントローラ13内のメモリが用いられる。

【0027】この後、ディスクアレイコントローラは、i番の不揮発性メモリが正常であるか否かのテストを行う(ステップS102)。i番の不揮発性メモリが正常でない場合(N)には、NERRに“1”を加算、Ndiskにiをセット(ステップS106)し、ステップS107に進む。

【0028】i番の不揮発性メモリが正常である場合(ステップS102; Y)には、その不揮発性メモリに管理情報C(i)をセット(ステップS103)する。C(i)の詳細については後述する。そして、i番のディスク装置の管理情報記録領域が正常に使用できるか否かの判断を行い(ステップS104)、管理情報記録領域に、異常が発見された場合(N)には、不揮発性メモリに異常が発見されたときと同じく、ステップS106に進む。

【0029】i番の管理情報記録領域が正常であると判断できたときには(ステップS104; Y)、そのディスク装置の管理情報記録領域にi番の不揮発性メモリの内容C(i)を書き込み(ステップS105)、ステップS107に進む。

【0030】ステップS107では、パラメータiに“1”が加算され、iが全ディスク装置の台数以下である場合(ステップS108; Y)には、ステップS102に戻り、次の不揮発性メモリと管理情報記録領域のテストが行われる。iが全ディスク台数より大きくなったとき(ステップS108; N)に、NERRの値の判定(ステップS109)が行われる。

【0031】NERRが“0”である場合には、全ディスク装置が正常であるので、初期化動作を完了させる。NERRが“2”以上の値であるときには、ディスクアレイ装置の動作が不可能となるので、エラーメッセージをホスト計算機に出力(ステップS110)して、初期化動作を終了する。この場合、ディスクアレイ装置を記憶装置として動作させることはできない。また、NERRが“1”であるときには、データを冗長構成とすることはできないが、記憶装置として動作させることが可能であるので、Ndisk番のディスク装置を使用不可として(ステップS111)、初期化動作を終了する。なお、この使用不可の設定は、ディスクアレイコントローラ13内のメモリの所定領域に、Ndisk番のディスク装置が使用不可であることを示す情報を書き込むことにより行っている。

【0032】このディスクアレイ装置では、初期化時に不揮発性メモリおよび管理情報記録領域に書き込む管理

情報C(i)として、11バイトのディスクアレイ装置識別情報と1バイトのスロット番号と2バイトの日付情報と2バイトの時刻情報を有する16バイトの情報を用いた。日付と時刻情報は、時計回路16から得られる情報であり、管理情報C(i)中では、初期化指示がなされた日付と時刻を示す情報を用いている。管理情報は、これに限られるものではなく、たとえば、乱数を用いることや、スロット番号だけを用いることや、任意の数値を用いることもできる。しかし、ディスク装置ごとに、対応する管理情報が異なるようにすることが望ましい。

【0033】次に、記憶装置として動作しているときに、ディスク装置141~145のうちの1台に障害が発生した際の動作を説明する。ディスクアレイコントローラは、個々のディスク装置から読み出し、または、書き込み障害が発生したことを示す情報を受信すると、そのディスク装置に対応する不揮発性メモリの管理情報の変更を行う。このディスクアレイ装置では、管理情報中の日付、時刻情報に対応する部分のみが、時計回路の出力を基に、障害が発生した日付、時刻情報で書き換えられるように構成してある。管理情報として他の形式の情報をを用いた場合には、この変更方法もそれに応じて修正される。たとえば、スロット番号を管理情報として用いる場合には、障害が発生するごとに、管理情報にディスクアレイ装置が内蔵するディスク装置の台数と等しい値(または、それより大きな値)が加算されるように装置を構成すればよい。

【0034】実施例のディスクアレイ装置では、この不揮発性メモリの管理情報とディスク媒体上の記録領域に書き込まれている管理情報を基に、装置の再立ち上げが行われたときには、以下のような手順で、各ディスク装置の状態の判定を行う。

【0035】図3と図4に、再立ち上げ時のディスクアレイコントローラ13の動作の流れを示す。図4に示した動作が図3に示した動作に引き続いて行われる。

【0036】まず、配列変数A(i)、B(i)に“0”を、カウンタiに“1”、障害ディスク装置台数カウンタNERRに“0”、障害ディスク装置識別パラメータNdiskに“0”をセット(ステップS201)する。そして、i番の不揮発性メモリのテストを行い(ステップS202)、不揮発性メモリが正常でない場合(N)には、そのディスク装置を使用不可と判断して、NERRを“1”増加させるとともに、Ndiskにiをセット(ステップS207)し、ステップS208に進む。

【0037】不揮発性メモリが正常である場合(ステップS202; Y)には、その不揮発性メモリの内容を、配列変数A(i)に書き込む。次いで、i番のディスク装置の管理情報記録領域のテストを行い、不揮発性メモリ記憶部が正常でない場合(ステップS204; N)には、ステップS207へ進む。管理情報記録領域が正常である場合(ステップS204; Y)には、その管理情



報記録領域の内容をB(i)にセット(ステップS205)する。ディスク装置に障害が発生していても、ディスク装置のカウント記録部が読み出せる場合はあるが、前述のように、障害発生時には、不揮発性メモリの内容が変更されている。このため、配列変数A(i)とB(i)の一致を判定(ステップS206)することにより、i番のディスク装置が障害ディスク装置であるか否かの判断が行える。

【0038】配列変数A(i)とB(i)が一致していないときには(ステップS206;N)、i番のディスク装置は、正常なディスク装置ではないので、障害ディスク装置台数カウンタNERRと障害ディスク装置識別パラメータNDISKを変更(ステップS207)した後に、iに“1”を加算(ステップS208)し、iが全ディスク台数より大きくないときには(ステップS209;Y)、ステップS202へ戻り、次の不揮発性メモリ、管理情報記録領域のテストを行う。

【0039】このような動作をiが全ディスク台数より大きくなるまで、繰り返すことにより(ステップS209;N)、NERRには、正常でないディスク装置の台数がセットされ、NDISKには、正常でないディスク装置のうち1台の識別番号がセットされることになる。

【0040】図4を用いて、図3に引き続いて行われる動作を説明する。まず、障害ディスク装置台数カウンタNERRの判定を行い(ステップS301)、NERRが“0”であるときには、全ディスク装置が正常であるので、立ち上げ動作を終了する。NERRが“2”以上であるときには、実施例のディスクアレイ装置の冗長構成により、カバーできない障害であるため、エラーメッセージを出力(ステップS302)して、再立ち上げ動作を終了する。

【0041】NERRが“1”であるときには、A(NDISK)と“0”との比較を行い、A(NDISK)が“0”であるときには(ステップS303;Y)、図3に示した流れ図から明らかなように、NDISK番の不揮発性メモリが異常であるため、NDISK番のディスク装置を使用不可とする(ステップS306)。A(NDISK)が“0”でないときには(ステップS303;N)、NDISK番のディスク装置が使用可能な状態である場合(たとえば、正常なディスク装置と交換されている場合)が考えられるので、そのディスク装置のテストを行う(ステップS204)。そして、そのディスク装置を異常と判断したときには(ステップS205;N)、そのディスク装置を使用不可とし(ステップS306)、正常と判断したときには(ステップS305;Y)、そのディスク装置に他のディスク装置内のデータから再構築されたデータを書き込む復旧作業を行う(ステップS307)。復旧作業が完了したならば、そのディスク装置の管理情報記録領域に、A(NDISK)を書き込み、再立ち上げ動作を完了する。

【0042】以上説明した実施例のディスクアレイ装置では、管理情報記録領域に、障害発生時に変更された不揮発性メモリ内の管理情報を書き込むように構成してあるが、管理情報記録領域と不揮発性メモリ、両者の内容を、初期化時に用いた管理情報C(i)で書き換えるように装置を構成してもよい。また、不揮発性メモリを1つにして、管理情報を格納する領域を区別することにより、各ディスク装置に対応させてもよいことは当然である。

10 【0043】実施例のディスクアレイ装置では、1台のディスク装置の復旧しか行えない冗長構成を採用しているため、障害ディスク装置識別パラメータNDISKを1つしか持たず、また、障害ディスク装置台数カウンタNERRが2以上となるときには、エラーメッセージを出力するように構成してあるが、2台以上のディスク装置を復旧できる冗長構成を用いる場合には、障害ディスク装置識別パラメータを復旧可能な台数に応じて増やし、障害ディスク装置台数カウンタによる分岐条件を変更してやればよい。また、実施例のディスクアレイ装置は、図4に示したように、立ち上げ時に、自動的に復旧可能なディスク装置に対しては復旧を行うように構成してあるが、この機構を設けずに、ホスト計算機からの指示により、復旧作業が開始されるように構成してもよい。

【0044】

【発明の効果】以上説明したように、請求項1記載の発明によれば、各ディスク装置の状態を立ち上げ時に自動的に検出するため、システムダウン後の再立ち上げ時に誤動作を生ずることがない。

30 【0045】請求項2記載の発明によれば、各ディスク装置の状態を立ち上げ時に自動的に検出するため、システムダウン後の再立ち上げ時に誤動作を生ずることがない。また、ディスク装置の接続ミスをも検出するため、人為的ミスによりデータが失われることがない。

40 【0046】請求項3記載の発明によれば、各ディスク装置の状態を立ち上げ時に自動的に検出するため、システムダウン後の再立ち上げ時に誤動作を生ずることがない。また、ディスク装置の接続ミスをも検出するため、人為的ミスによりデータが失われることがない。さらに、障害ディスク装置が交換されたことを検出して自動的にその復旧動作を開始するため、復旧動作の開始指示をホスト計算機を介して入力する必要がない。

【図面の簡単な説明】

【図1】 本発明の実施例のディスクアレイ装置の概要を示す構成図である。

【図2】 実施例のディスクアレイ装置の初期化動作の流れを示す流れ図である。

【図3】 実施例のディスクアレイ装置の立ち上げ時における各ディスク装置の状態判定を行う動作の流れを示す流れ図である。

50 【図4】 実施例のディスクアレイ装置の立ち上げ時



13

14

に、図 3 の流れに引き続いて行われる動作の流れを示す流れ図である。

【図 5】 従来例のディスクアレイ装置の概要を示す構成図である。

【図 6】 従来例のディスクアレイ装置の書き込み動作の概要を示す説明図である。

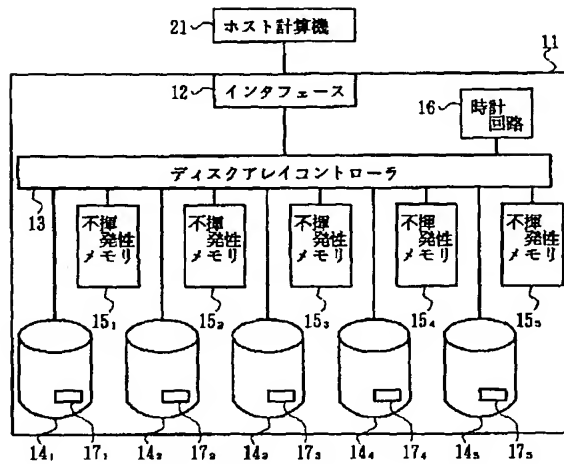
【図 7】 従来例の不揮発性記憶部を設けた 2 重化ディスク装置の概要を示す構成図である。

【図 8】 従来例のヒストリカウンタ部を設けた 2 重化ディスク装置の概要を示す構成図である。

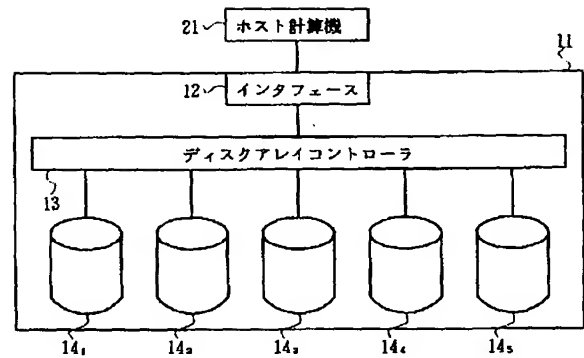
# 【符号の説明】

11…ディスクアレイ装置、12…インタフェース、13…ディスクアレイコントローラ、14、22…ディスク装置、15…不揮発性メモリ、16…時計回路、17…管理情報記録領域、18…記録領域、20…書き込み要求データ、21…ホスト計算機、23…制御部、24…不揮発性記憶部、25…オペレーティングシステム、26…チャネル装置、27…中央処理装置、28…主メモリ、29…ヒストリカウンタ部、30…ヒストリカウンタ検証部、31…制御テーブル

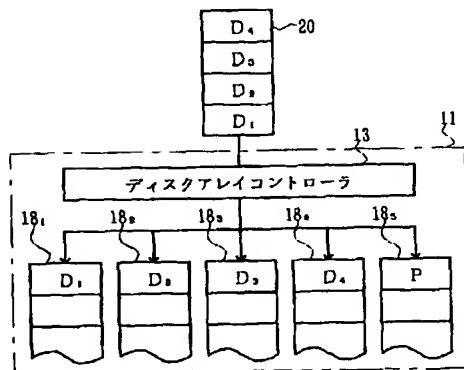
【図 1】



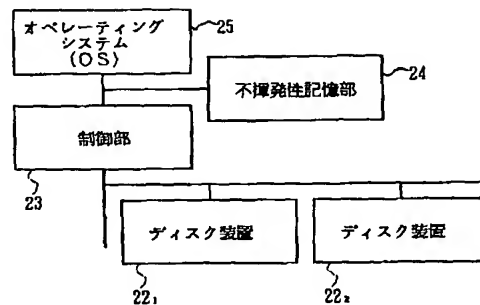
【図 5】



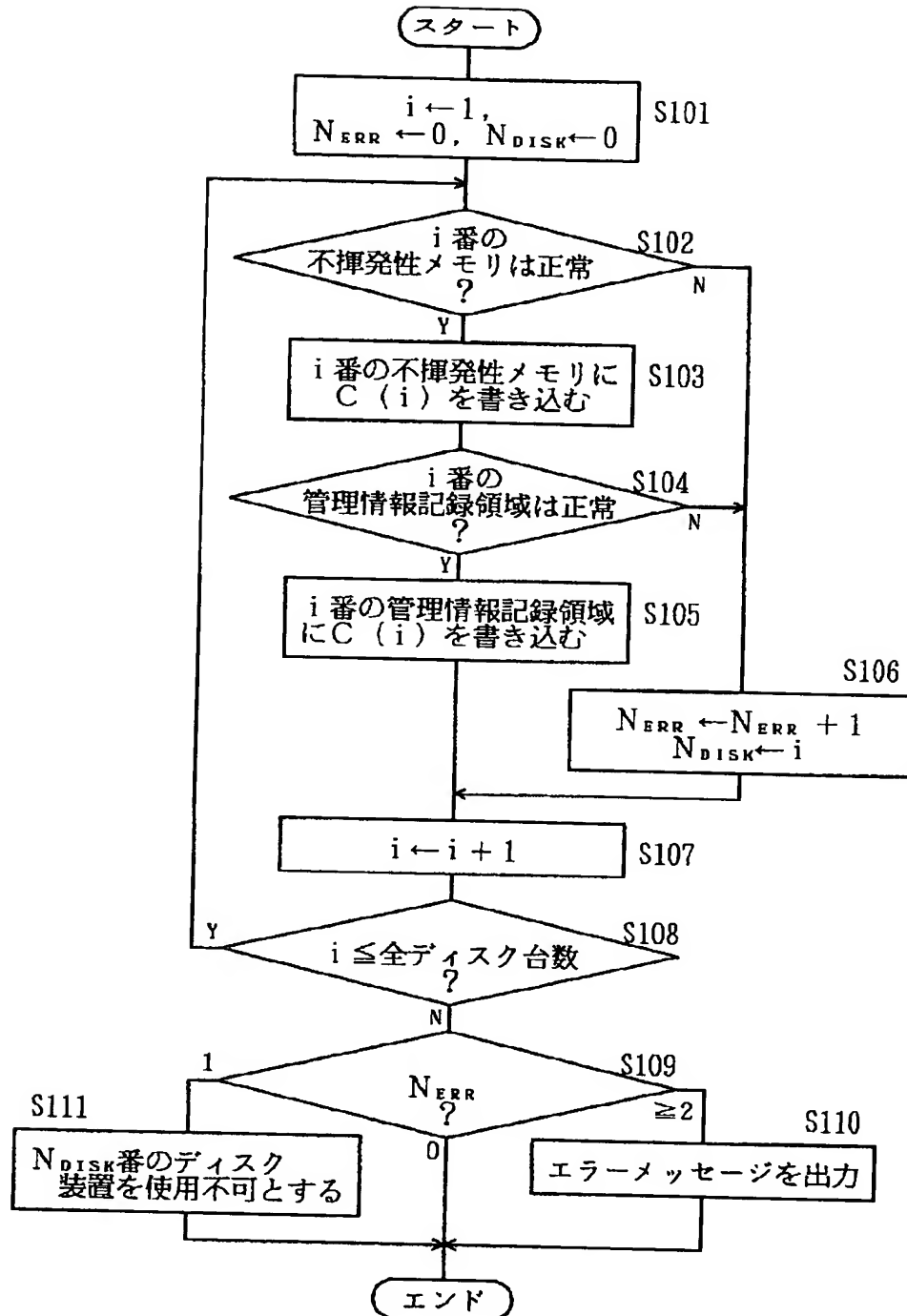
【図 6】



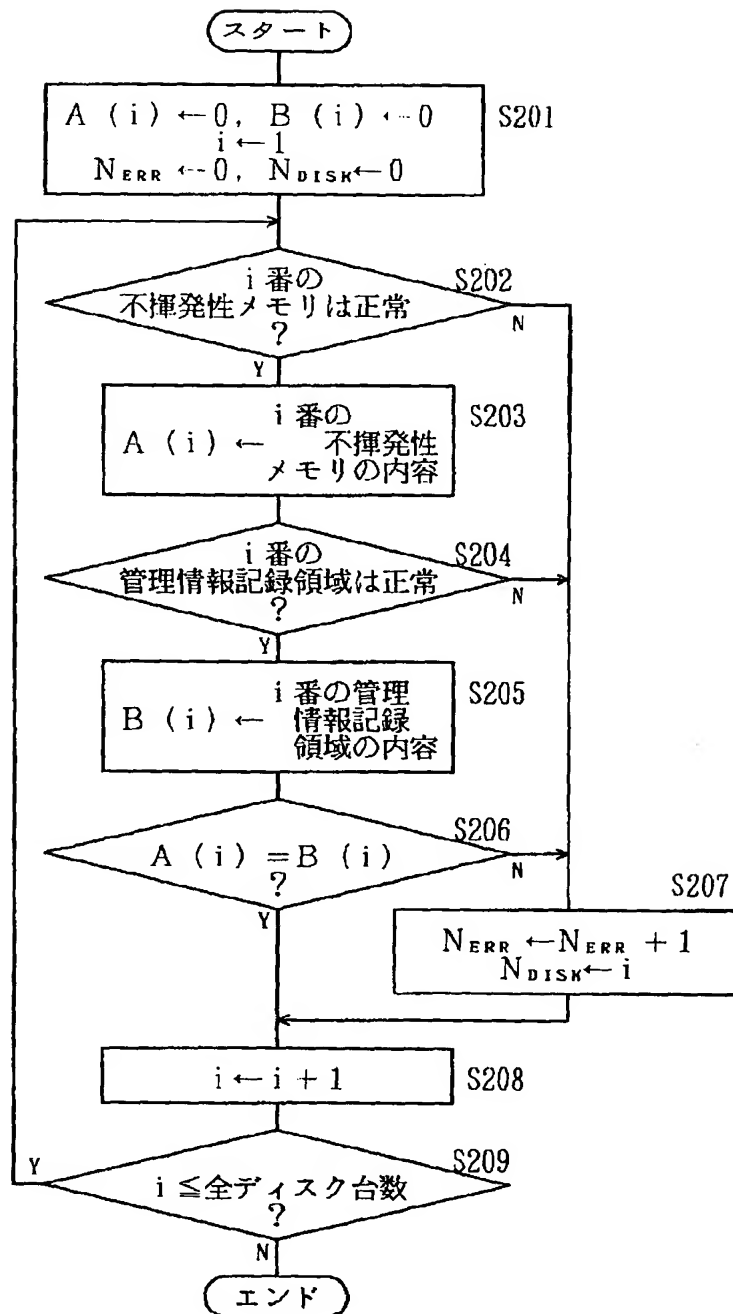
【図 7】



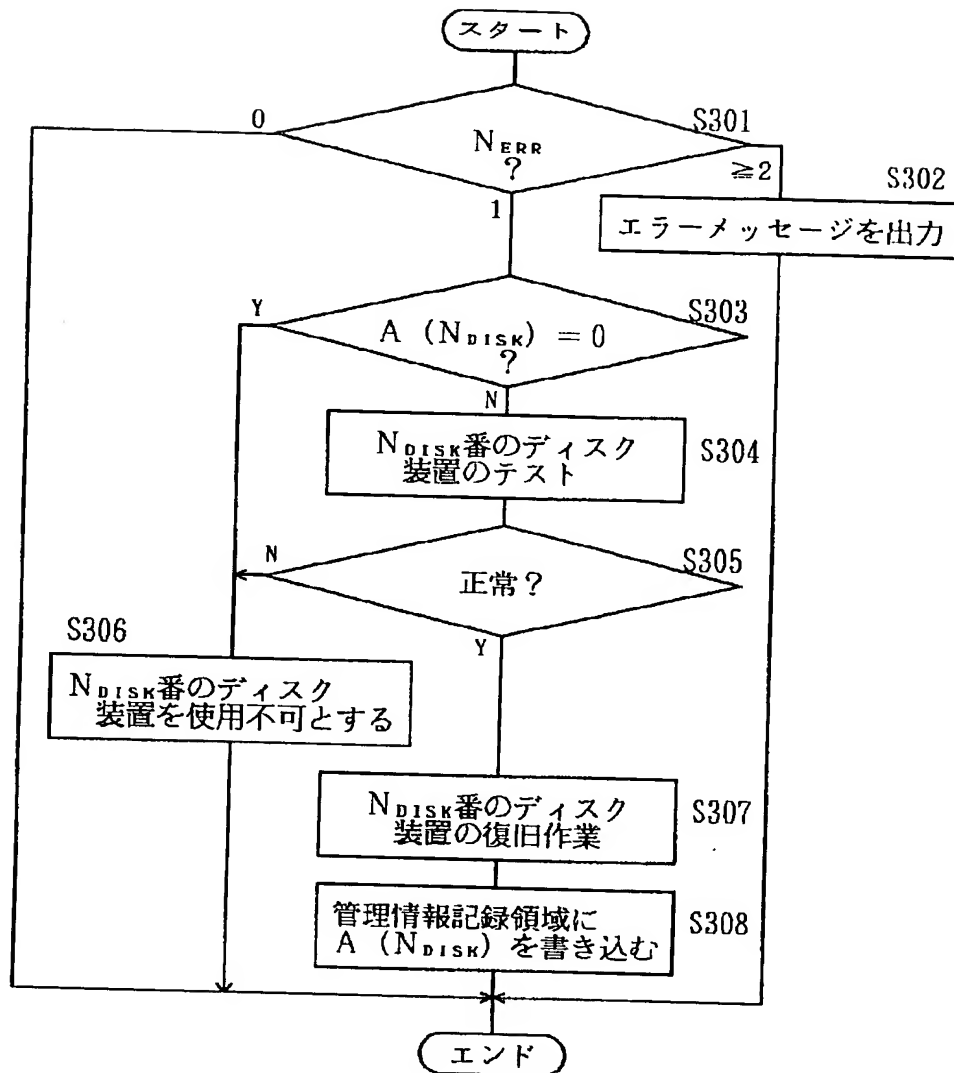
【図 2】



【図3】



【図4】



【図8】

